

106715 U.S. PRO
06/12/00

06/12/00
09/592572
JC856 U.S. PTO

Assistant Commissioner for Patents
Washington, D.C. 20231

Enclosed for filing is the utility patent application of RICHARD DELLAONA for HIGH SPEED INFORMATION PROCESSING AND MASS STORAGE SYSTEM AND METHOD, PARTICULARLY FOR INFORMATION AND APPLICATION SERVERS .

[x] 13 sheet(s) of [] formal [x] informal drawing(s);

☐ a claim for foreign priority under 35 U.S.C. §§ 119 and/or 365 is ☐ hereby made to __
 filed in __ on __;
☐ in the declaration;

[] a certified copy of the priority document;

☐ a General Authorization for Petitions for Extensions of Time and Payment of Fees;

[] _____ statement(s) claiming small entity status;

[] an Assignment document;

[] an Information Disclosure Statement; and

[] Other: _____

[x] An [] executed [x] unexecuted declaration of the inventor(s)
[] also is enclosed [] will follow.

[] Please amend the specification by inserting before the first line the sentence --This application claims priority under 35 U.S.C. §§119 and/or 365 to __ filed in __ on __; the entire content of which is hereby incorporated by reference.--

☐ A bibliographic data entry sheet is enclosed.



21839

☒ The filing fee has been calculated as follows ☐ and in accordance with the enclosed preliminary amendment:

CLAIMS					
	NO. OF CLAIMS		EXTRA CLAIMS	RATE	FEE
Basic Application Fee					\$690.00 (101)
Total Claims	36	MINUS 20 =	16	x \$18.00 (103)	288.00
Independent Claims	2	MINUS 3 =	0	x \$78.00 (102)	0.00
If multiple dependent claims are presented, add \$260.00 (104)					
Total Application Fee					978.00
If verified Statement claiming small entity status is enclosed, subtract 50% of Total Application Fee					
Add Assignment Recording Fee if Assignment document is enclosed					
TOTAL APPLICATION FEE DUE					978.00

☒ This application is being filed without a filing fee. Issuance of a Notice to File Missing Parts of Application is respectfully requested.

☒ This paper is submitted in duplicate.

Please address all correspondence concerning the present application to:

Frederick G. Michaud, Jr.
Burns, Doane, Swecker & Mathis, L.L.P.
P.O. Box 1404
Alexandria, Virginia 22313-1404.

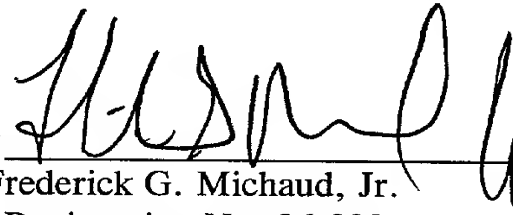
Respectfully submitted,

BURNS, DOANE, SWECKER & MATHIS, L.L.P.

Date:

6/12/00

By:



Frederick G. Michaud, Jr.
Registration No. 26,003

P.O. Box 1404
Alexandria, Virginia 22313-1404
(703) 836-6620

**HIGH SPEED INFORMATION PROCESSING AND MASS
STORAGE SYSTEM AND METHOD, PARTICULARLY FOR
INFORMATION AND APPLICATION SERVERS**

Related Applications:

This application is related to Patent Application No. PCT/US99/05231 of Richard Dellacona filed March 10, 1999, which is based upon Provisional Patent Application Serial No. 60/077,643, filed March 10, 1998, and to U.S. Patent Application Serial No. 09/071282 of Richard Dellacona filed May 1, 1998, all of which are hereby incorporated herein by reference.

BACKGROUND OF THE INVENTION

Field of the Invention

This invention relates to a high speed, microcomputer based, Fibre Channel compatible and fault tolerant information processing and mass storage system especially suited for information servers and application servers. In particular, the present invention relates to a method and apparatus for information processing and storage involving a unique and extremely versatile system architecture, including a dual loop arbitrated, Fibre Channel capable, multiple-fault tolerant, hot-swappable mass storage disk array and including a method and apparatus for providing an enterprise-wide information or application server system using such disk array.

State of the Art

Efforts have been made in the past to provide a mass storage file server capable of delivering information throughout an enterprise with high speed data throughput, scalable data storage capability in a convenient, easily configurable enclosure using well known, industry standard operating software and hardware. However, such systems have typically experienced many shortcomings and problems associated with equipment incompatibility as well as with the inability of presently available computer and communications hardware to sustain performance and service failure of component devices. Such shortcomings have included the lack of capability to add storage devices to accommodate increased storage requirements

or to replace failed storage devices without the need to completely power down the information server. Some of the compatibility problems have involved, for example, bottlenecks in sharing information among the equipment components of various vendors. The above-referenced Dellacona patent applications address some of these
5 problems and others, and provide unique solutions which are described and claimed therein.

The present invention further addresses some of the problems discussed in the referenced Dellacona applications, as well as others. For example, the present invention further addresses the problem of scalability and customization in
10 information processing and storage systems for different applications. In some applications, there may be a greater need for processing capability rather than storage capacity, while other applications may require just the opposite. Yet other applications may require storage expansion for existing information processing systems. This invention provides an architecture which will readily accommodate
15 such needs.

In addition, mass storage systems can create considerable heat, particularly where they are disk drive based. If the heat is not effectively removed, it can affect the reliability and life of the system. Often, it is difficult to remove the heat because of obstructions caused by the physical configuration of back planes and mid planes
20 which act as barriers to air flow. Typically, for example, all of the disk drives of a mass storage module or array are typically plugged into connectors on the face of a backplane or mid plane that extends across the entire module. Whether enclosed in a cabinet or other enclosure or rack mounted without an enclosure, air flow through the module is inhibited by this type of structural arrangement, and excessive heating
25 can occur, particularly in the vicinity of the disk drives.

Also, it is desirable to be able to hot swap individual disk drives of a mass storage module to accommodate the need for more storage capacity, but the system storage requirements may outgrow the capacity of the module and it may also be desirable to have the capability of adding modules without powering down the
30 system. Of course, this capability must be provided without disturbing the operation of the existing module and with a minimum of signal degradation as modules are added.

SUMMARY OF THE INVENTION

The present invention obviates one or more of the foregoing problems and/or shortcomings of the prior art through the provision of an information processing and mass storage method and system, including a unique mass storage array, particularly
5 suited for information servers or application servers, with a novel system architecture which permits the addition or replacement of storage devices without interrupting or seriously degrading the operation of the system and which is highly fault-tolerant and reliable. In addition, the invention obviates one or more of the foregoing problems by providing a novel physical layout that permits the effective
10 removal of heat from a system module containing heat creating components.

In accordance with one embodiment of the invention, an information processing and mass storage system adapted to be readily expandable to increase its storage capacity while the system is in operation comprises at least one module containing (a) at least one computer, (b) a plurality of plug-in storage devices such
15 as disk drives for storing information, (c) a storage device bypass circuit board associated with each storage device, with each storage device being plugged into a connector on the bypass circuit board, (d) a module bypass circuit board including an optical input/ output connector for outputting electrical signals from the module as light signals and for inputting light signals into the module as electrical signals
20 and (e) a controller connecting the computer with each of the storage devices through its associated storage device bypass circuit board and through the module bypass circuit board.

Certain information server configurations in accordance with the invention include one or more computers, each computer connected to communicate through a
25 Fibre Channel controller with a mass storage array comprising a plurality of bypass circuit boards, at least some of which are connected to an information storage device. In one embodiment, the controller provides a dual loop communications channel comprising two complete communication paths to each of bypass circuit boards and associated storage devices. In another embodiment, the controller
30 provides a single loop which traverses the bypass circuit boards and any associated storage devices twice. In a one configuration with two or more computers, the Fibre

Channel controller connected to each computer communicates with the mass storage array through a Fibre Channel controller bypass card.

In a preferred embodiment the computer is preferably a suitable conventional single board computer. The controller preferably is a conventional arbitrated dual
5 channel Fibre Channel system through which the computer communicates with each of the storage device bypass circuit boards and the module bypass circuit board. The bypass circuit boards may be any suitable circuits which form a continuous loop for the Fiber Channel controller regardless of whether a disk drive is plugged into the drive bypass circuit board. The loop continues through other modules when they are
10 connected to the module bypass circuit thereby readily permitting expansion while maintaining a unitary information processing and mass storage system.

In accordance with another embodiment of the present invention, a high speed information processing and mass storage system includes two modules, each including a plurality of disk drives in a hot-swappable, disk drive array. Each disk
15 drive array is connected to a module bypass circuit which includes an optical input/output connector, preferably an optoelectronic transceiver. The optical input/output connectors of the modules are connected by a fiber optic transmission medium such that signals are communicated between the modules in the form of light. With this configuration, modules may be added to increase storage capacity
20 without interrupting the operation of each other and without serious signal degradation.

In accordance with yet another embodiment of the present invention, a high speed information processing and/or mass storage system with disk drives for information storage includes at least one module with a plurality of drive bypass
25 circuit boards, each including a drive bypass circuit board connector. At least one opening is provided between connectors to permit the flow of air between the connectors. Each drive bypass circuit board is a relatively flat circuit board with connectors on different edges of the board, wherein one of the connectors is the connector which receives the disk drive and the other connector connects to said
30 drive bypass circuit board connector. The connectors, bypass circuit boards and drives are arranged such that when they are connected there is a path for air flow from outside the module alongside each bypass circuit board and its associated disk

drive for cooling purposes without any backplane obstruction. Where the mass storage system is housed in an enclosure, at least one fan is mounted to force air from outside said enclosure through the spaces between said bypass boards and drives, preferably through the openings between the drive bypass circuit board connectors.

It will be appreciated that the present invention provides a novel high speed information processing and/or mass storage system particularly suitable for information and application servers. The system is scalable, fault tolerant, and reliable both because of its novel system architecture and its physical layout. Other features and advantages of the invention will become apparent from the following detailed description of exemplary and preferred embodiments when read in conjunction with the drawings which illustrate, by way of example, the principles of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a functional block diagram generally illustrating an information or application server system according to the present invention using a high speed mass storage system of the invention;

Figures 2A, 2B and 2C are diagrammatic representations illustrating various system configurations for different applications wherein different ratios of processing and storage are required;

Figure 3 is a functional block diagram illustrating a mass storage module of Figure 1 in greater detail;

Figure 4A is a functional block diagram illustrating the dual loop communication path from the Fibre Channel controller of Figure 3 connected so as to provide one logical Fibre Channel communication path traversing each of the storage device bypass circuit boards twice;

Figure 4B is a functional block diagram illustrating the dual loop communication path from the Fibre Channel controller of Figure 3 connected so as to provide two logical Fibre Channel communication paths traversing each of the storage device bypass circuit boards once, thereby being independently available to communicate with each of the storage devices;

Figure 5 is a functional block diagram illustrating an information or application server system according to the present invention wherein two or more servers are connected to a mass storage array which may include one or more mass storage modules;

5 Figure 6 is a functional block diagram illustrating an embodiment of a web server application configured in accordance with the architecture of the present invention;

10 Figure 7 is a functional block diagram illustrating an embodiment of a basic video streaming application configured in accordance with the architecture of the present invention;

Figure 8 is a functional block diagram illustrating another embodiment of a video streaming application configured in accordance with the architecture of the present invention and including a duplicated index server;

15 Figure 9 is a functional block diagram illustrating another embodiment of a video streaming application configured in accordance with the architecture of the present invention and including a distributed index server;

Figure 10 is a functional block diagram illustrating one embodiment of a storage device bypass circuit board according to the present invention;

20 Figure 11 is a functional block diagram illustrating one embodiment of a chassis bypass circuit board according to the present invention;

Figures 12 A and 12 B are diagrammatic representations of the physical layout of a drive bypass circuit board and disk drive illustrating the preferred connector arrangement according to the present invention; and

25 Figures 13 A and 13 B are diagrammatic representations of a connector arrangement for connecting a single board computer (SBC) to various input/out devices according to the present invention.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

30 One embodiment of an information or application server system in accordance with the present invention using the high speed information processing and mass storage system of the invention is illustrated in Figure 1. Referring to Figure 1, the server, designated generally by the reference numeral 100, includes a computer 102, a controller 104, preferably a Fibre Channel controller, and a

communications interface or access card 106. The server 100 communicates via the Fibre Channel controller 104 with a mass storage array 200 which includes one or more mass storage modules 202A ... 202n. As illustrated, the computer 102 may also communicate with a suitable diagnostic computer 110 as is described in the
5 aforementioned Dellacona applications.

The computer 102 preferably comprises single board high speed computer running a computer industry standard operating system software program such as, for example, Windows NT available from Microsoft Corporation. An operating system like Windows NT may be stripped down to remove those elements of the
10 program which are not needed, if desired to preserve memory or to increase operating speed. Suitable conventional drivers of the type used for similar applications may be provided as necessary to support the particular architecture being implemented.

The computer may include a display such as a touch screen display, and
15 various storage and peripheral devices (not shown) as required. The single board computer can include any of a wide number of suitable devices including, but not limited to, the Compact PCI CPU Board with Pentium Processor, Model No. ZT 5510, available from Ziatech Corporation. Modifications to enhance performance of the ZT 5510 can include an onboard 40 MB flash memory card for permanent
20 storage of the non-reconfigurable portions of the Windows NT operating system software and an onboard, removable, PCMCIA 40 Mb flash memory card, "D2 FlashDisk" available from Sandisk Corporation for read/writeable storage of the reconfigurable portions of the Windows NT software.

The Fibre Channel controller 104 may be any suitable design according to
25 the Fibre Channel Consortium created as a separate board or incorporated into the single board computer design. The communications interface or access card 106 may be any suitable device made in accordance with well known T-1 communications architecture, and/or architecture adapted for compatibility with other network and telecommunications architectures, protocols and topologies
30 including, but not limited to, T-3, DS-3, OC-3C, OC-12C, OC-192C, FDDI, SONET, SCSI, TCP/IP, HiPPI and ATM. In addition, the computers 102 and 110 may be networked together and with other computers through appropriate ethernet

cards or other suitable networking techniques. The respective manufacturer, Fibre Channel Consortium and I2O Special Interest Group reference design data sheets and materials describing the detailed operating capabilities and specifications of each of the foregoing components are hereby incorporated by reference in their entirety. Also, further information concerning possible subsystems and connection protocols for the server are described in the referenced Dellacona applications.

Figures 2A, 2B and 2C illustrate generally three types of system configurations which can be addressed in accordance with the architecture of the present invention. Each Figure is diagrammatically illustrative of a chassis 300 with a power supply section 302 and a diagnostics section 304. In addition, the chassis includes a processing section 306 such one or more of the servers 100 of Figure 1, and /or a storage section 308 such as one or more of the mass storage modules 202 of Figure 1. The size of the processing and storage sections can be readily configured for a particular application.

For example, Figure 2A illustrates an arrangement for a high volume server such as an application server, a movie server (e.g., for video streaming to multiple users) or communications in conjunction with a carrier-class switch. It can be seen that the processing section 306 includes ten slots, each preferably representing a single board computer. In contrast, the storage section 308 has only five storage slots, each representing a high speed, high capacity storage device such as a disk drive. Thus, the processing function is stressed over the storage function. On the other hand, Figure 2B illustrates an arrangement which might be suitable for a web-server or web-hosting. Here, the processing section 306 comprises two slots of the chassis whereas the storage section 308 comprises fifteen slots. Figure 2C illustrates a chassis arrangement suitable primarily for storage expansion. Here, there is no processing section and the chassis is devoted to storage.

Figure 3 illustrates an embodiment of the mass storage array 200 of Figure 1 in greater detail as it could be configured with a single computer server 100 and one or more mass storage modules 202. Referring to Figure 2, the mass storage array 200 includes one or more modules 202A, 202B ... 202n, each of which is preferably identical architecturally, although they may contain different numbers of storage devices and different types of storage devices. Each module 202 includes bypass

circuit boards 210A, 210B ... 210n and at least some storage devices 212A, 212B 212n connected to the boards 210. In addition, a each storage device is preferably provided with an associated read/write switch 214A, 214B ... 214n, respectively.

Each module 202 has a module or chassis bypass circuit board 216 in the
5 communication path of the Fibre Channel controller 104. The chassis bypass circuit board 216 of module 202A is provided with an optical input/ output connector 218 for outputting electrical signals from the module 202A as light signals and for inputting light signals into the module 202A as electrical signals. Likewise, the modules 202B .. 202n have chassis bypass circuit boards 216 with associated
10 input/output connectors 218 (not shown). As illustrated, the input/output connector 218 of the chassis bypass circuit board of module 202A is adapted to be connected via a light transmission medium such as optical fibers 220 to the optical input/output connector 218 (not shown) of the chassis bypass circuit board of module 202B.

As was previously noted, the controller 104 preferably is a conventional
15 Fibre Channel Controller (FCC) which operates on a Fibre Channel protocol, and preferably is an arbitrated dual channel Fiber Channel Controller. The controller 104 provides a dual channel communication path within each module 202 between the computer 102 and each of the operable storage devices 212. As is described hereinafter in greater detail, the bypass circuit boards 210 ensure that the
20 communication path is complete even if a storage device 212 is inoperable (i.e., is not operable at or above some minimum level as is hereinafter described in greater detail) or has been removed from the connector on the bypass circuit board.

In that regard, each of the storage devices 212 is preferably a high speed, high capacity, conventionally available disk drive which is removably connected to
25 its associated bypass circuit board 210. Preferably, the disk drive plugs into a connector on the bypass circuit board 210 so that it can be readily removed and replaced or so that drives may be added to empty bypass circuit board positions, as needed to expand the storage capacity of the module. Each bypass circuit board 210 includes circuits which connect the controller 104 to the disk drive 212 when the
30 disk drive is plugged in and is conveying to the bypass circuit board that it is operable at a certain minimum level. On the other hand, the bypass circuit board 210 connects the controller 104 directly to the next bypass circuit board, bypassing

the disk drive 212, when the disk drive is not plugged in or is not operating at or above the minimum level. Any suitable, conventional disk drive of the type that runs self-diagnostics and provides a fault/no fault output signal may be used for this purpose.

5 With further reference to Figure 3, by way of example, the bypass circuit board 210A directs communications to the associated storage device 212A and then to the next bypass circuit board 210B when an operable storage device 212A is connected to the bypass circuit board 210A. When the storage device 212A is inoperable, i.e. not operating at some minimum satisfactory level, or when there is
10 no storage device connected to the bypass circuit board 210A, e.g., if the storage device is removed for replacement, the dual channel communication path proceeds through the bypass circuit board 210A without interruption, i.e., bypasses the storage device connector. While one embodiment of a bypass circuit board to accomplish the foregoing connection and bypass functions is described hereinafter in greater
15 detail, it will be appreciated by one skilled in the art that these functions can be accomplished in any suitable conventional manner by electronic switching circuits controlled by the fault/no fault signal from the storage device.

 With continued reference to Figure 3, the module or chassis bypass circuit board 216 provides functions similar to the storage or drive bypass circuit boards
20 210 in the sense that they either route communications back to the controller directly if there is no additional module 202B connected to the module 202A or they route communications to the next module 202B if one is connected and signals are being received. In the case of a single module 202, therefore, there is a dual channel communications loop entirely within the module by which the computer 102
25 communicates with each of the disk drives 212. On the other hand, by virtue of the chassis bypass circuit board 216, the dual channel communications loop extends to each of the disk drives 212 in the next module when one is connected. Specifically, if there is an additional module 202B connected to module 202A as illustrated, then communications from the computer 102 of module 202A are directed by the module
30 bypass circuit board 216 to the next module 202B via connectors 218 and optical fibers 220 so that module 202B is within and traversed by the dual Fibre Channel communication loop.

It will be appreciated that with the above described architecture, individual mass storage device modules 202 of the mass storage array 200 may be expanded internally by adding disk drives or other suitable storage devices, and bad disk drives may be replaced without affecting the operation of the rest of the module or the system it is used in. This provides an extremely versatile hot swappable, hot expandable, mass storage device array. In addition, when the demands of the system exceed the capacity of a single module, an additional module may be added, again without interrupting the operation of the rest of the array or its system.

As was explained above, the Fibre Channel controller provides a dual path through each module of the mass storage array 200. In accordance with the present invention, the system can be configured so that the dual path is a single path which traverses the mass storage array twice or two independent paths as is illustrated in Figures 4A and 4B.

Referring now to Figures 4A and 4B, the Fibre Channel controller (FCC) 104 includes two output paths A and B. Each path traverses the mass storage module 202 as illustrated, communicating with each of the present and operable storage devices 212 and returning to the FCC. In the Figure 4A embodiment, the A path returning to the FCC is looped back to the B output path. Accordingly, a single continuous path traverses the mass storage module twice. In the Figure 4B embodiment, the A return path is not looped back to the B output path. Accordingly, two paths traverse the mass storage module and can be used independently. This latter embodiment provides a second path in the event that one path fails.

It will be appreciated by one skilled in the art that the Figure 4A and 4B embodiments provide redundancy and fault tolerance. The Figure 4A embodiment is somewhat simpler to implement because only one set of chips is necessary to provide the single Fibre Channel capability required for the single loop. Still, if one of the loops is broken or encounters some other fault, that loop can be bypassed within the controller, and the other loop is still available. Similarly, while the Figure 4B embodiment may be more complex and expensive to implement, it provides two independently addressable loops for fault tolerance and redundancy, but also provides significantly greater communications bandwidth.

Figure 5 illustrates a further configuration which is made possible by the system architecture of the present invention. In the Figure 5 embodiment, two servers are connected to a mass storage array, a first module of which is illustrated. The servers 100A and 100B are connected to Fibre Channel bypass boards 222A and 222B, respectively. The operation of the computers 102A and 102B may be sensed by the Fibre Channel bypass boards, for example by sensing signal flow as with the chassis bypass board, so that the computers 102A and 102B can operate together or, in the case of a fault, separately. As was previously mentioned, each server may have flash memory, and in the embodiment of Figure 5, each computer 102A and 102B has its own flash memory with its operating system stored therein. In this fashion, the computers can boot independently from its associated flash memory rather than from the shared memory or other disk arrangement.

It will be appreciated that the system configuration illustrated in Figure 5 provides two servers and thus increased processing power. In addition, one server provides backup to the other in the event of failure, thereby providing increased fault tolerance and reliability. In addition, the architecture of Figure 5 permits the two servers to communicate with each other at Fibre Channel speeds, higher than could be achieved with 100BaseT LAN, using IP protocol.

It can be seen that system architecture according to the present invention lends itself to a wide variety of configurations to accommodate a variety of applications. Figure 6 illustrates one configuration wherein two servers A and B are connected in a suitable local area network (LAN) configuration with access to the internet. Server A has its own mass storage array 200 as does server B. It will be appreciated that several servers and storage arrays can be networked in this fashion to provide a very powerful information or application server system particularly suitable for web server applications. In addition, this configuration, permits full duplication of computer/mass storage systems rather than sharing a mass storage system as with the embodiment of Figure 5. Moreover, while the Figure 6 embodiment illustrates communication over a LAN which typically may be a 100BaseT LAN, it will be appreciated that this may be a Fibre Channel LAN if rates in the gigabit range are desired. These features and advantages may be the ideal choice for critical processing applications such as web servers.

Figure 7 illustrates another configuration particularly useful for video streaming applications. In this embodiment, the diagnostics computer 110 also has indexing functions for controlling access to content servers 1, 2 and 3. The content servers 100 provide access to video stored in their associated storage arrays 200.

5 Alternatives, also particularly suitable for video streaming applications, are shown in Figures 8 and 9. In the Figure 8 embodiment duplicated index servers separate from the diagnostics computer are provided with their own storage arrays. In the illustrated embodiment there are two index servers 1 and 2 and there are three content servers 1, 2 and 3, each with an associated storage array. In the Figure 9 embodiment, the functions of the index servers are distributed among the content
10 servers so that there are multiple index/content servers 1-5. One skilled in the art will appreciate that the Figure 8 approach uses an architecture where the index server or servers coordinate content streaming from the content servers whereas in Figure 9 the index and content server functionality is distributed across all servers,
15 providing extensive scalability.

Figures 10 and 11 are functional block diagrams which generally illustrate the storage and chassis bypass circuit boards 210 and 216, respectively, in greater detail. Referring to Figure 10, the storage bypass circuit board 210 includes a bypass board backplane connector 230 arranged to plug into a connector on the
20 backplane generally indicated at 232. The A and B signal paths coming from a previous bypass circuit or directly from the Fibre Channel controller are connected through the backplane to the bypass board via the backplane connector 230. Similarly, the A and B signal paths emerge from the bypass board 210 via the bypass board backplane connector 230.

25 The A signal path from the backplane connector 230 is connected to a suitable conventional electronic switch 234. The B signal path from the backplane connector 230 likewise is connected to an electronic switch 236. The A and B signal paths from electronic switches 234 and 236 are connected to a bypass board storage card or drive connector 238 where they are routed to the storage device (e.g., a disk
30 drive) 212.

The return A signal path from the bypass board drive connector 238 is connected to the switch 234, and the return B signal path from the connector 238 is

connected to switch 236. A fault signal produced by the storage device to indicate its presence and its level of operability as was described above is applied to each of the electronic switches 234 and 236 to control the switching thereof. The A and B return paths from the switches 234 and 236 are connected to the bypass board
5 backplane connector 230 where they are routed through the backplane 232 to the next bypass circuit board or to the Fibre Channel controller.

In operation, the A signal path enters the bypass circuit board and is connected to the switch 234. If the fault signal is not present (i.e., there is no fault and the signal is in a low or negative signal state) indicating that the storage device
10 is not present or is inoperable, the switch 234 returns the A signal path to the bypass board backplane connector 230 thus bypassing the storage device 212. The B signal path similarly is looped back to the backplane connector 230 by the electronic switch 236 if the fault signal is low. On the other hand, the A and B signal paths are routed through the switches 234 and 236 to the storage device and then back through the
15 switches when the fault signal is high or positive indicating the storage device is present and operable.

Referring now to Figure 11, the chassis bypass circuit board is essentially the same as the storage bypass circuit board except the selection made by the electronic switches 234 and 236 is between acting as a bypass or connecting the A and B to the
20 I/O connector 238. In this regard, the I/O connector is preferably a conventional optical fiber transceiver for use in bi-directional communication applications over multimode optical fiber, particularly in multimode or single mode Fibre Channel applications. For example, the transceiver may be a model MLC-25-6-X-T optical fibre channel small factor (SFF) transceiver available from Methode Electronics, Inc.
25 of Chicago, Illinois. Such transceivers include a light transmitter and receiver as well as a standard receptacle for receiving an industry standard optical fiber connector. In addition, the transceiver provides a signal detect output (the fault signal in Figure 11) which indicates whether or not the transceiver is receiving a light signal. If it is not, the fault signal causes the electronic switches 234A and
30 236A to loop the A and B paths back to the chassis bypass board connector 230A. If, on the other hand, the fault signal indicates that a light signal is being received by the transceiver 218, the electrical signals on the A and B paths are passed through

the switches 234A and 236A to the transceiver 218 where they are converted to light signals and transmitted over the optical fibers forming the A and B paths to the next chassis. Similarly, light signals returned from the next chassis on the A and B paths are converted back to electrical signals by the transceiver 218 and returned along the A and B paths through the switches 234A and 236A to the connector 230A and onto the next bypass circuit or the Fibre Channel controller.

Typically, each of the modules 202 of a mass storage device array such as the one shown in Figure 3 is enclosed in a housing or, if mounted in a rack, may be surrounded by other structures and/or other circuit boards. The server components such as the computers 102 and 110, the touch screen display 112, the Fibre Channel controller, and other system components such as communication access cards, ethernet cards, LAN components and the like may also be mounted within the same housing or on the same rack. Heating may therefore be a problem, particularly where the storage devices used in the modules are disk drives driven by motors. Accordingly, the preferred embodiment of the present invention as illustrated in Figures 12A and 12B includes a physical structure and arrangement of the various bypass circuit boards and storage devices that accommodates the circulation of cooling air throughout the module without minimal obstruction.

In accordance with one aspect of this physical structure of the invention, each of the drive bypass circuit boards 210 is a relatively thin circuit board. The circuit board is, however, unlike typical circuit boards built to receive a disk drive or other mass storage device. In such conventional circuit boards, the connector or plug which receives the plug-in disk drive typically is positioned so that the disk drive extends perpendicular to the plane of the board. For example, it is usual to have the bypass circuits and/or the communications paths between them on a backplane or midplane circuit board which extends across the module in a fashion similar to a computer motherboard which extends across the computer chassis and has connectors to receive various plug-in cards or boards. That arrangement creates an obstruction which makes it more difficult to effectively cool heat producing storage devices such as disk drives.

As is illustrated in Figures 12A and 12B, the present invention does not use either a circuit board midplane or a backplane structure. Instead, the connector on

the drive bypass board which receives the disk drive (e.g., the connector 238 in Figure 10) is at the edge of the board so that the drive, when plugged in, extends parallel to the circuit board. In addition, the drive bypass circuit board itself is not plugged into a backplane circuit board. Rather, a connector is provided on the edge of the board, preferably opposite the disk drive connector as illustrated at 239 in Figure 10, and that connector plugs into an individual connector which is mounted on a frame or other structural member of the chassis and which is wired or otherwise connected to similar connectors for the other disk bypass circuit boards.

Figures 12A shows a side view of the disk drive 212 plugged into the drive bypass circuit board 210 via connector 238, with the drive bypass circuit board plugged into connector 240 suitably mounted on structural members 242 of the chassis or rack containing the mass storage module and/or server and its associated components. Figure 12 B is an end view illustrating several bypass circuit boards 21 plugged into the connectors 240 which are in turn connected by screws or other suitable means to the structural members 242. Since there is no backplane which would normally make the connections between adjacent components, electrical or light connections generally indicated at 246 are suitably provided between the connectors 240 to provide the communications required, as illustrated, for example, in Figure 3 and 5.

It can be seen that between each connector 240 there is a space 244 through which air can readily be drawn or forced by a fan or other air circulation means as is necessary. Even if the structural members are part of a housing that surrounds the components, screening or other suitable openings can be provided so that the areas 244 permit sufficient air flow. It can also be seen that because the disk drive is plugged into the bypass circuit board so that their planes are parallel and not perpendicular, there is no obstruction of air flow.

It will also be appreciated that this arrangement is particularly suited for field service of the unit and is readily upgradeable. Connectors can be readily replaced in the field without the need to change a complete backplane or midplane board, and in some instances repairs of this sort can be carried out with little or no down time. In addition, the illustrated connection arrangement permits expansion without the limitations encountered when using a backplane or midplane with a fixed number of

expansion slots and without the other physical and electrical limitations encountered with backplanes or midplanes.

Similar physical arrangements may be used to connect computers to their associated components to create the desired server configuration. For example, as
5 illustrated in Figure 13A, a single board computer of the type previously described in connection with the description of the server 100 may connect to the communications card 106 through a mini compactPCI (CPCI) backplane. In this case, the single board computer SBC may incorporate the Fibre Channel controller FCC. Likewise, in Figure 13B, the SBC may connect to an input/output unit I/O at
10 the left rear connector of the CPCI backplane. In this case, the I/O may include the FCC functions. Again, it will be appreciated that the foregoing advantages are achieved with this sort of simple backplane or midplane structure which does not extend across the cabinet or rack.

The above-described exemplary embodiments are intended to be illustrative
15 in all respects, rather than restrictive, of the present invention. Thus the present invention is capable of many variations in detailed implementation that can be derived from the description contained herein by a person skilled in the art. All such variations and modifications are considered to be within the scope and spirit of the present invention as defined by the following claims.

WHAT IS CLAIMED IS:

1 1. A high speed mass storage system which is readily expandable to
2 increase its storage capacity while the system is in operation comprising at least one
3 module containing

4 (a) at least one CPU;

5 (b) a plurality of plug-in storage devices for
6 storing information;

7 (c) a storage device bypass circuit board
8 associated with each storage device, each storage device being
9 plugged into a connector on the storage device bypass circuit
10 board;

11 (d) a module bypass circuit board including an
12 optical input/ output connector for outputting electrical signals
13 from the module as light signals and for inputting light signals
14 into the module as electrical signals; and,

15 (e) a controller providing a communication path
16 between the CPU with each of the storage devices through its
17 associated storage device bypass circuit board and through the
18 module bypass circuit board.

1 2. The high speed mass storage system of claim 1 wherein each storage
2 device bypass circuit board includes a circuit which completes the connection of the
3 CPU with the other storage device bypass circuits and their associated storage
4 devices whether or not the storage device is present.

1 3. The high speed mass storage system of claim 1 wherein the module
2 bypass circuit board outputs electrical signals from the at least one module via the
3 optical input/output connector when light signals are received by said optical
4 input/output connector.

1 4. The high speed mass storage system of claim 1 including first and
2 second modules each including elements (a) through (e), wherein the optical
3 input/output connectors of the modules are connected by a fiber optic transmission
4 medium such that signals are communicated between the modules in the form of
5 light.

1 5. The high speed mass storage system of claim 4 wherein the module
2 bypass circuit board of the first module outputs electrical signals from the first
3 module to the second module via the optical input/output connector when light
4 signals are received from the second module by said optical input/output connector.

1 6. The high speed mass storage system of claim 1 wherein the at least
2 one module includes a storage device bypass board connector for each of the storage
3 device bypass circuit boards with an opening between each connector to permit the
4 flow of air between the connectors and alongside the bypass circuit boards and
5 storage devices for cooling purposes.

1 7. The high speed mass storage system of claim 1 wherein the storage
2 devices are disk drives and the storage device bypass circuit boards are disk drive
3 bypass circuit boards each having a connector to receive a disk drive.

1 8. The high speed mass storage system of claim 7 wherein the at least
2 one module includes a disk drive bypass circuit board connector for each of the drive
3 bypass circuit boards with an opening between each connector to permit the flow of
4 air between the connectors, and wherein each drive bypass circuit board is a
5 relatively flat circuit board with a connector on opposite edges, wherein one of the
6 connectors is the connector which receives the disk drive and the other connector
7 connects to said drive bypass circuit board connector, said connectors, bypass circuit
8 boards and drives being arranged such that when they are connected there is a path
9 for air flow from outside the module alongside each bypass circuit board and its
10 associated disk drive for cooling purposes without any backplane obstruction.

9. The high speed mass storage system of claim 8 wherein the module is housed in an enclosure and at least one fan is mounted to force air from outside said enclosure through the spaces between said bypass boards and drives.

10. The high speed mass storage system of claim 1 wherein the controller operates with a Fiber Channel protocol.

11. The high speed mass storage system of claim 1 wherein the controller is an arbitrated dual channel Fiber Channel controller.

12. The high speed mass storage system of claim 10 wherein each storage device is a disk drive and wherein each storage device bypass circuit board comprises a disk drive bypass circuit board including a circuit which completes the connection of the CPU with the other drive bypass circuits and their associated disk drives whether or not the disk drive is present.

13. The high speed mass storage system of claim 12 wherein the module bypass circuit board outputs electrical signals from the at least one module via the optical input/output connector when light signals are received by said optical input/output connector.

14. The high speed mass storage system of claim 11 including first and second modules each including elements (a) through (e), wherein the optical input/output connectors of the modules are connected by a fiber optic transmission medium such that signals are communicated between the modules in the form of light.

15. The high speed mass storage system of claim 14 wherein the module bypass circuit board of the first module outputs electrical signals from the first module to the second module via the optical input/output connector when light signals are received from the second module by said optical input/output connector.

1 16. The high speed mass storage system of claim 11 wherein the storage
2 devices are disk drives and the storage device bypass circuit boards are disk drive
3 bypass circuit boards, and wherein the at least one module includes a disk drive
4 bypass board connector for each of the disk drive bypass circuit boards with an
5 opening between each connector to permit the flow of air between the connectors
6 and alongside the bypass circuit boards and disk drives for cooling purposes.

1 17. The high speed mass storage system of claim 12 wherein the at least
2 one module includes a disk drive bypass circuit board connector for each of the disk
3 drive bypass circuit boards with an opening between each connector to permit the
4 flow of air between the connectors, and wherein each disk drive bypass circuit board
5 is a relatively flat circuit board with a connector on opposite edges, wherein one of
6 the connectors is the connector which receives the disk drive and the other connector
7 connects to said disk drive bypass circuit board connector, said connectors, bypass
8 circuit boards and drives being arranged such that when they are connected there is a
9 path for air flow from outside the module alongside each bypass circuit board and its
10 associated disk drive for cooling purposes without any backplane obstruction.

1 18. The high speed mass storage system of claim 17 wherein the module
2 is housed in an enclosure and at least one fan is mounted to force air from outside
3 said enclosure through the spaces between said bypass boards and drives.

1 19. A high speed mass storage system adapted to be readily expandable
2 to increase its capacity while the system is in operation comprising at least one
3 module containing

- 4 (a) at least one CPU;
- 5 (b) a plurality of plug-in disk drives for storing
6 information;
- 7 (c) a disk drive bypass circuit board associated
8 with each disk drive and including a disk drive connector at
9 one edge thereof and a bypass board connector at another edge

10 thereof, each disk drive being plugged into said disk drive
11 connector on the disk drive bypass circuit board;

12 (d) a module bypass circuit board including an
13 optical input/ output connector for outputting electrical signals
14 from the module as light signals and for inputting light signals
15 into the module as electrical signals; and,

16 (e) a controller connecting the CPU with each of
17 the disk drives through its associated drive bypass circuit
18 board and through the module bypass circuit board such that a
19 loop is formed between the output and input of the controller
20 with each disk drive bypass circuit board and the module
21 bypass circuit board in said loop and completing said loop
22 whether or not a disk drive is plugged into the disk drive
23 connector.

1 20. The high speed mass storage system of claim 19 including a second
2 module connected to said at least one module via the optical input/output connector,
3 said module bypass circuit board of the at least one module completing said loop
4 through the second module.

1 21. The high speed mass storage system of claim 19 wherein the module
2 bypass circuit board outputs electrical signals from the at least one module via the
3 optical input/output connector when light signals are received by said optical
4 input/output connector.

1 22. The high speed mass storage system of claim 20 wherein each of said
2 modules includes elements (a) through (e), wherein the optical input/output
3 connectors of the modules are connected by a fiber optic transmission medium such
4 that signals are communicated between the modules in the form of light.

1 23. The high speed mass storage system of claim 22 wherein the module
2 bypass circuit board of the at least one module outputs electrical signals from the at

3 least one module to the second module via the optical input/output connector when
4 light signals are received from the second module by said optical input/output
5 connector .

1 24. The high speed mass storage system of claim 19 wherein the at least
2 one module includes a drive bypass board connector for each of the drive bypass
3 circuit boards with an opening between each connector to permit the flow of air
4 between the connectors and alongside the bypass circuit boards and disk drives for
5 cooling purposes.

1 25. The high speed mass storage system of claim 19 wherein the at least
2 one module includes a drive bypass circuit board connector for each of the drive
3 bypass circuit boards, and wherein each drive bypass circuit board is a relatively flat
4 circuit board with a connector on opposite edges, wherein one of the connectors is
5 the connector which receives the disk drive and the other connector connects to said
6 drive bypass circuit board connector, said connectors, bypass circuit boards and
7 drives being arranged such that when they are connected there is a path for air flow
8 from outside the module alongside each bypass circuit board and its associated disk
9 drive for cooling purposes without any backplane obstruction.

1 26. The high speed mass storage system of claim 25 wherein the at least
2 one module is housed in an enclosure and at least one fan is mounted to force air
3 from outside said enclosure through the spaces between said bypass boards and
4 drives.

1 27. The high speed mass storage system of claim 26 wherein each drive
2 bypass circuit board connector is mounted in the same plane in spaced relationship
3 with each other.

1 28. The high speed mass storage system of claim 19 wherein the
2 controller operates with a Fiber Channel protocol.

1 29. The high speed mass storage system of claim 19 wherein the
2 controller is an arbitrated dual channel Fiber Channel system.

1 30. The high speed mass storage system of claim 29 wherein each drive
2 bypass circuit board includes a circuit which completes the connection of the CPU
3 with the other drive bypass circuits and their associated disk drives whether or not
4 the disk drive is present.

1 31. The high speed mass storage system of claim 29 wherein the module
2 bypass circuit board outputs electrical signals from the at least one module via the
3 optical input/output connector when light signals are received by said optical
4 input/output connector.

1 32. The high speed mass storage system of claim 29 including first and
2 second modules each including elements (a) through (e), wherein the optical
3 input/output connectors of the modules are connected by a fiber optic transmission
4 medium such that signals are communicated between the modules in the form of
5 light.

1 33. The high speed mass storage system of claim 32 wherein the module
2 bypass circuit board of the first module outputs electrical signals from the first
3 module to the second module via the optical input/output connector when light
4 signals are received from the second module by said optical input/output connector.

1 34. The high speed mass storage system of claim 29 wherein the at least
2 one module includes a drive bypass board connector for each of the drive bypass
3 circuit boards with an opening between each connector to permit the flow of air
4 between the connectors and alongside the bypass circuit boards and disk drives for
5 cooling purposes.

1 35. The high speed mass storage system of claim 29 wherein the at least
2 one module includes a drive bypass circuit board connector for each of the drive

3 bypass circuit boards with an opening between each connector to permit the flow of
4 air between the connectors, and wherein each drive bypass circuit board is a
5 relatively flat circuit board with a connector on opposite edges, wherein one of the
6 connectors is the connector which receives the disk drive and the other connector
7 connects to said drive bypass circuit board connector, said connectors, bypass circuit
8 boards and drives being arranged such that when they are connected there is a path
9 for air flow from outside the module alongside each bypass circuit board and its
10 associated disk drive for cooling purposes without any backplane obstruction.

1 36. The high speed mass storage system of claim 35 wherein the module
2 is housed in an enclosure and at least one fan is mounted to force air from outside
3 said enclosure through the spaces between said bypass boards and drives.

ABSTRACT OF THE DISCLOSURE

A high speed, microcomputer based, Fibre Channel compatible and fault tolerant information processing and mass storage system especially suited for information servers and application servers. A unique and extremely versatile system architecture, including a dual loop arbitrated, Fibre Channel capable, multiple-fault tolerant, hot-swappable mass storage disk array, permits combinations of servers and mass storage arrays which can be tailored for a wide variety of applications and which can be configured with emphasis on the system characteristics such as redundancy, speed, processing capability, storage capability, and the like, as desired. A unique backplane and/or midplane arrangement for connecting the system components allows for easy and, in most cases, on-line field upgrading and/or service and at the same time provides for the very effective cooling of components, particularly those such as disk drives which tend to produce a lot of heat.

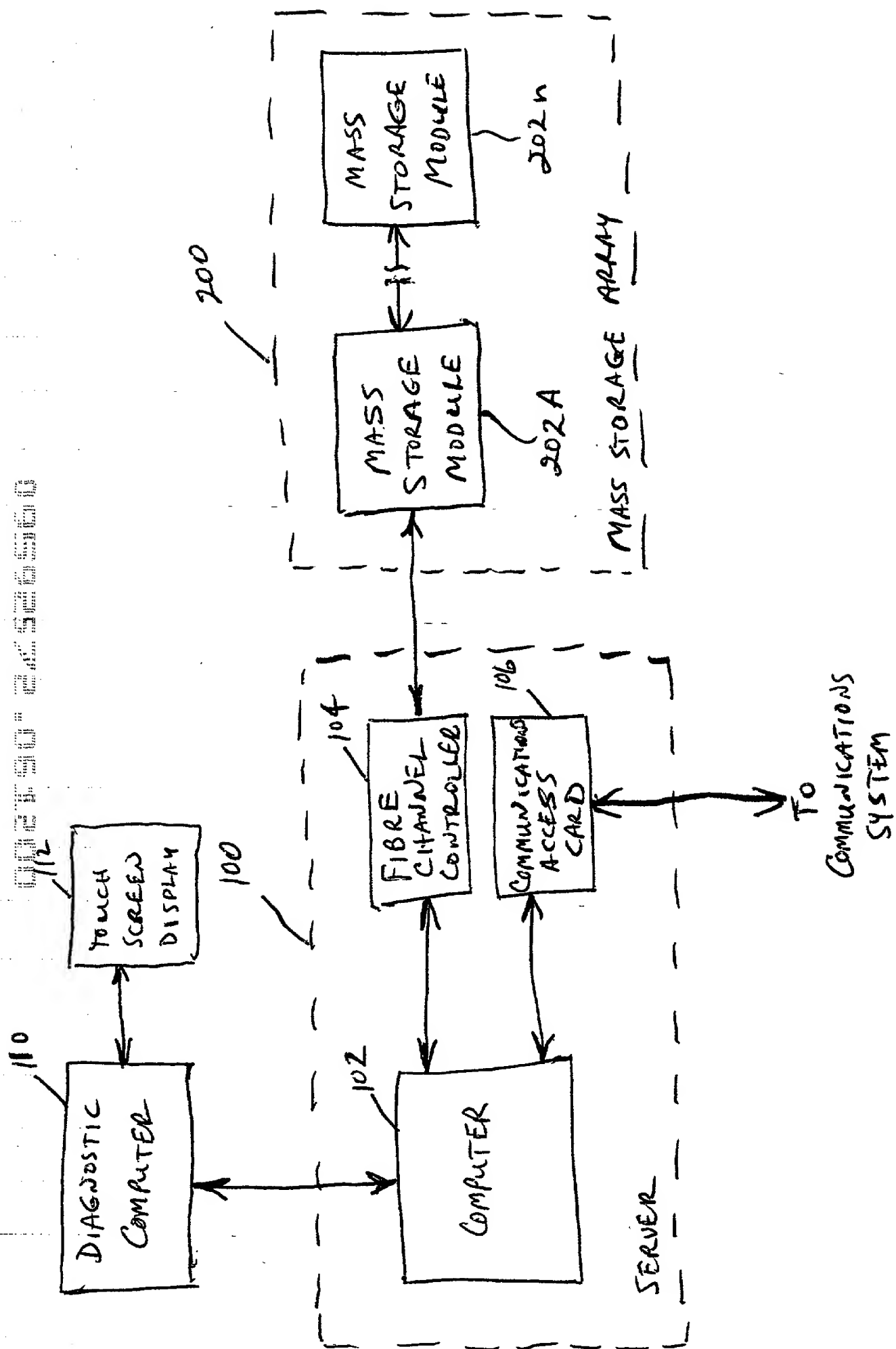


FIG. 1

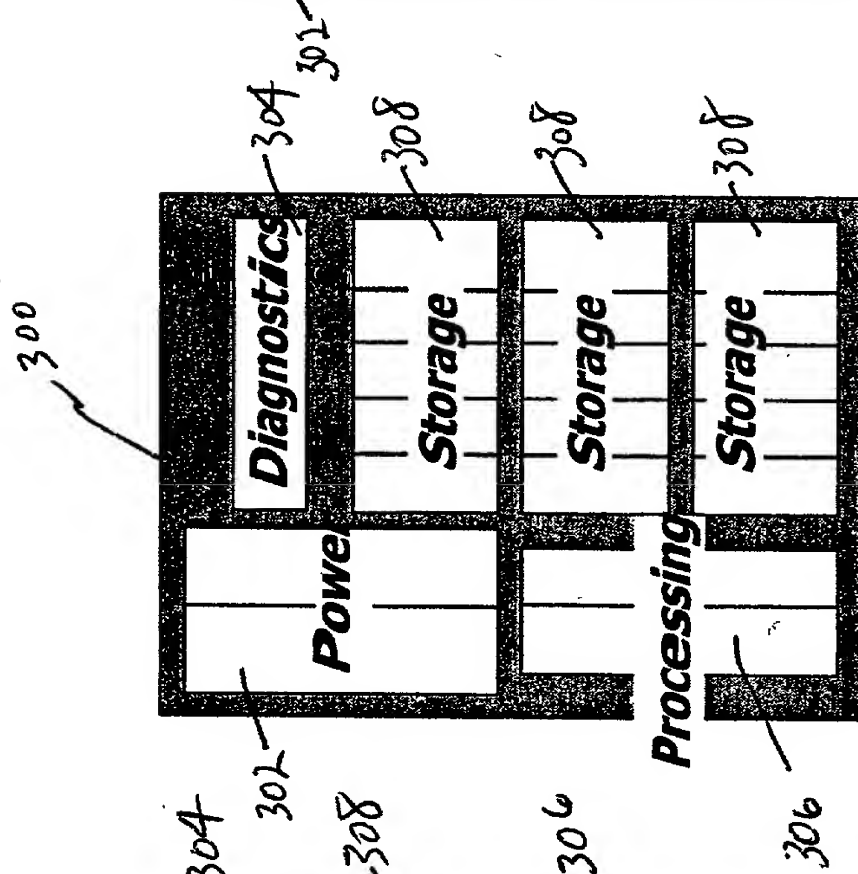
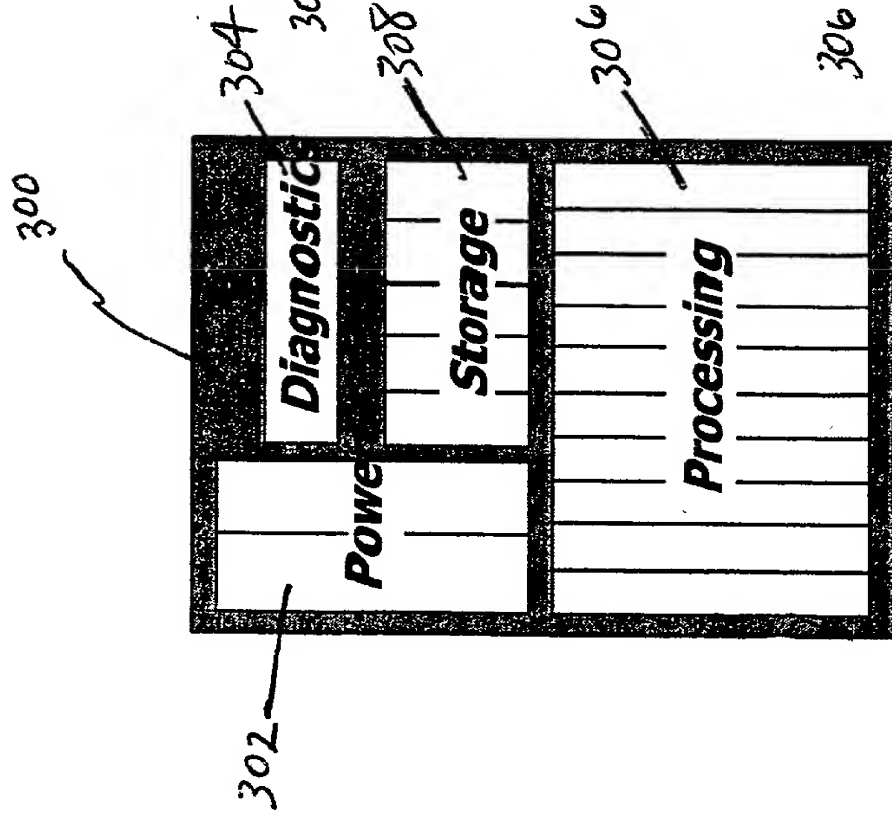


FIG. 2B

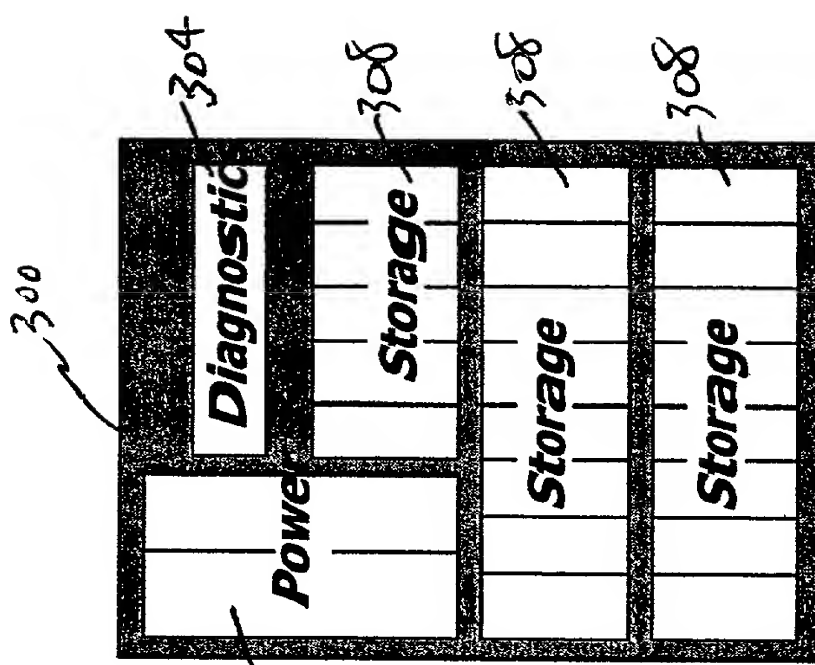
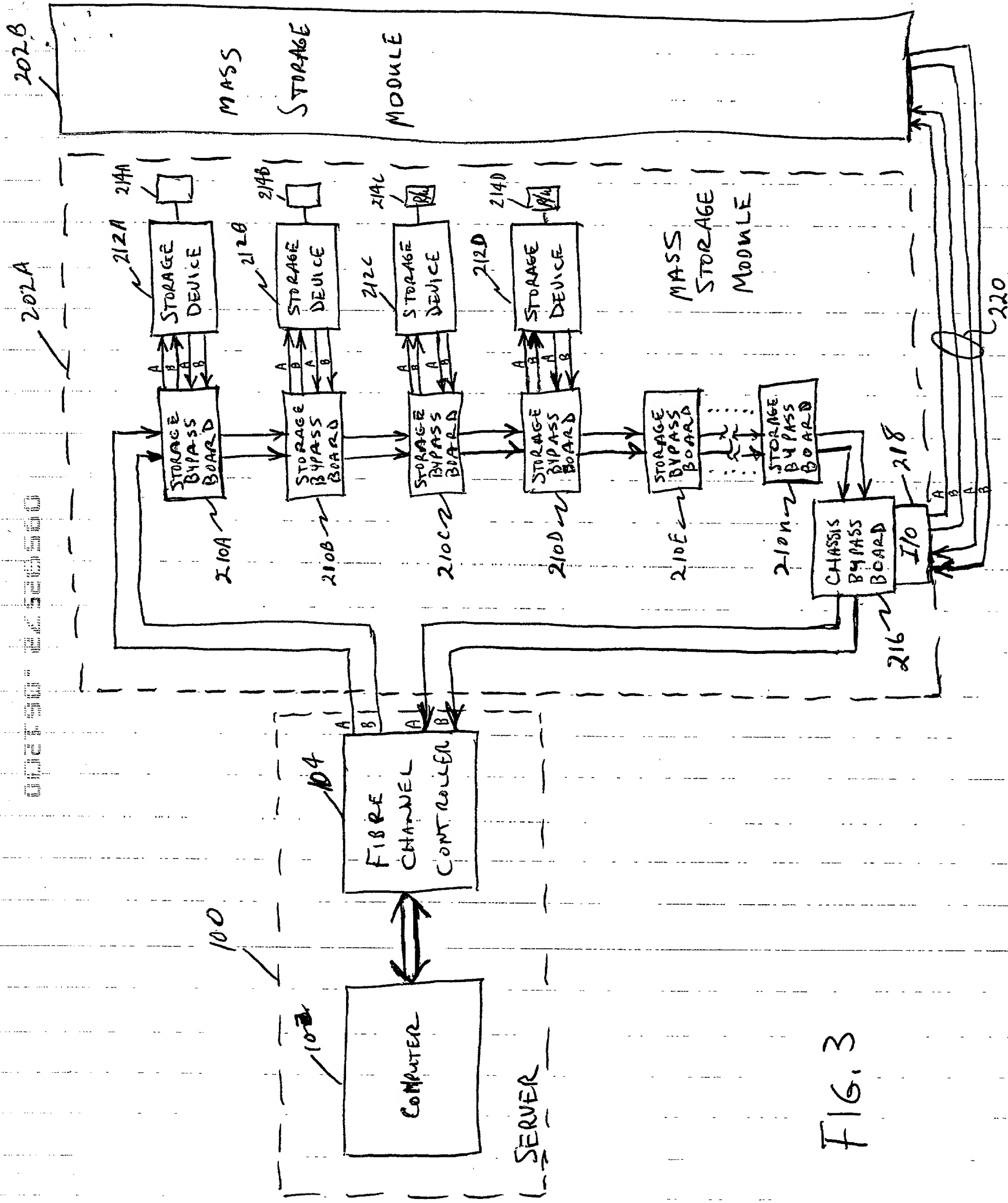


FIG. 2C



361

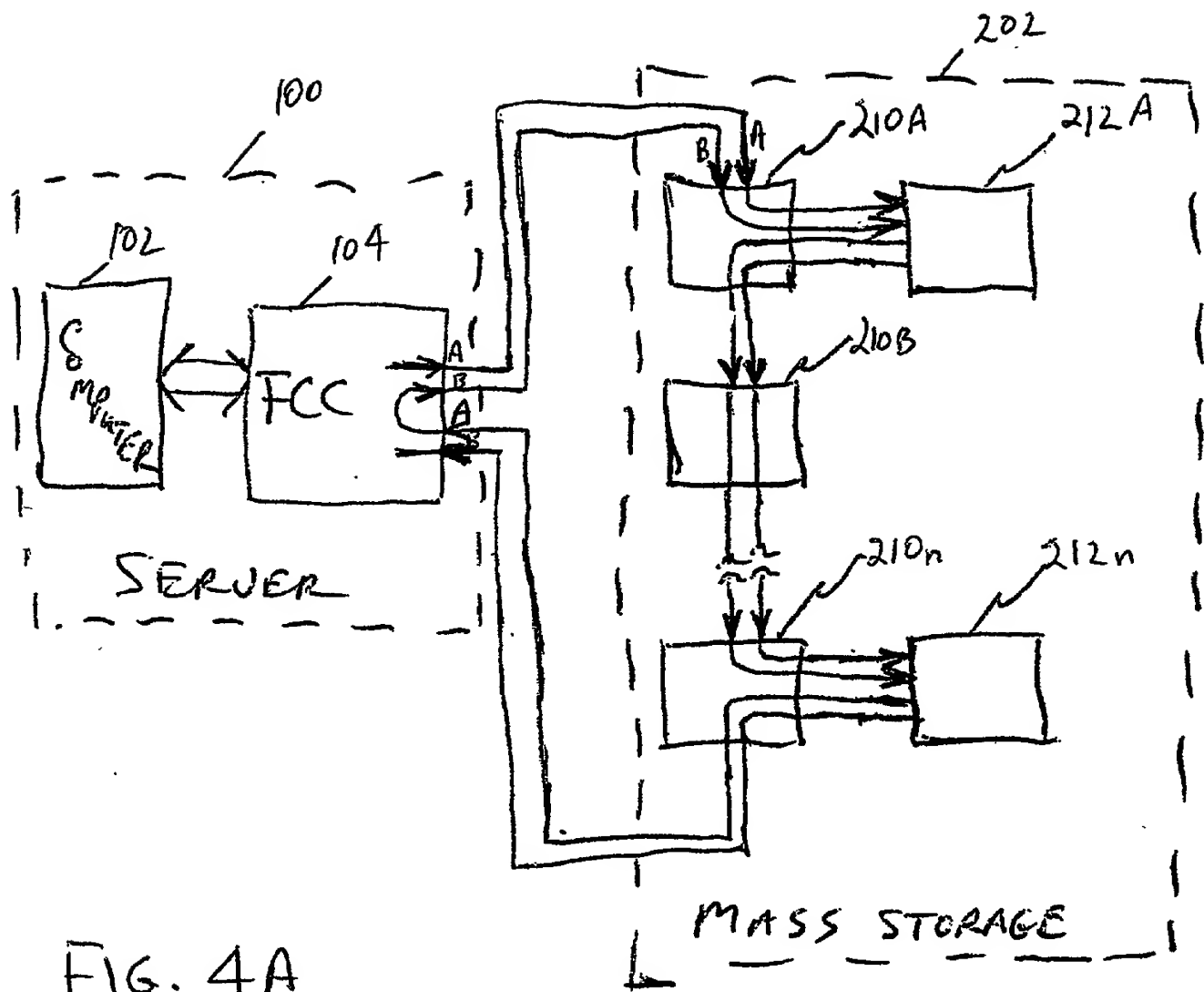


FIG. 4A

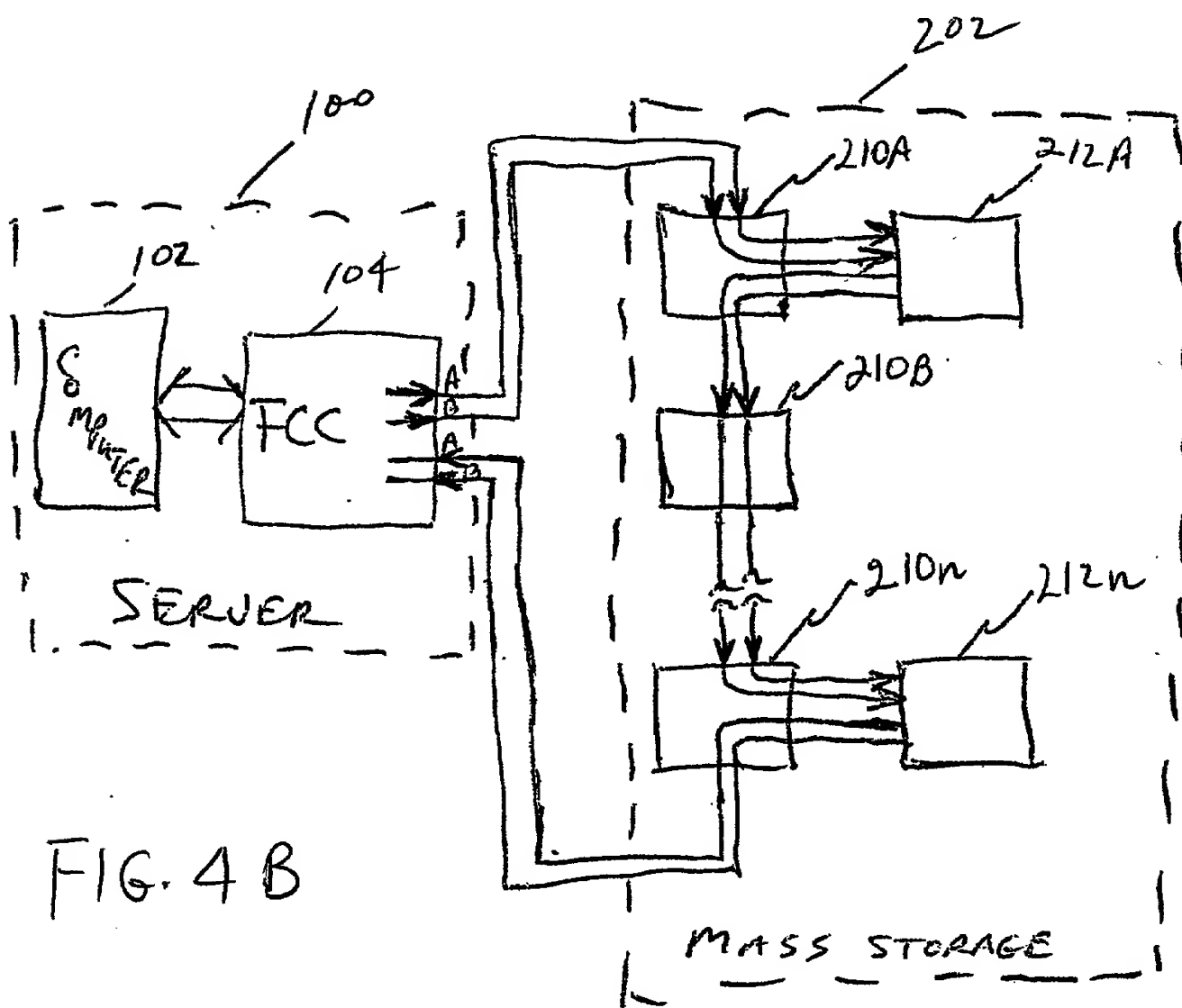


FIG. 4B

FIG. 5

202

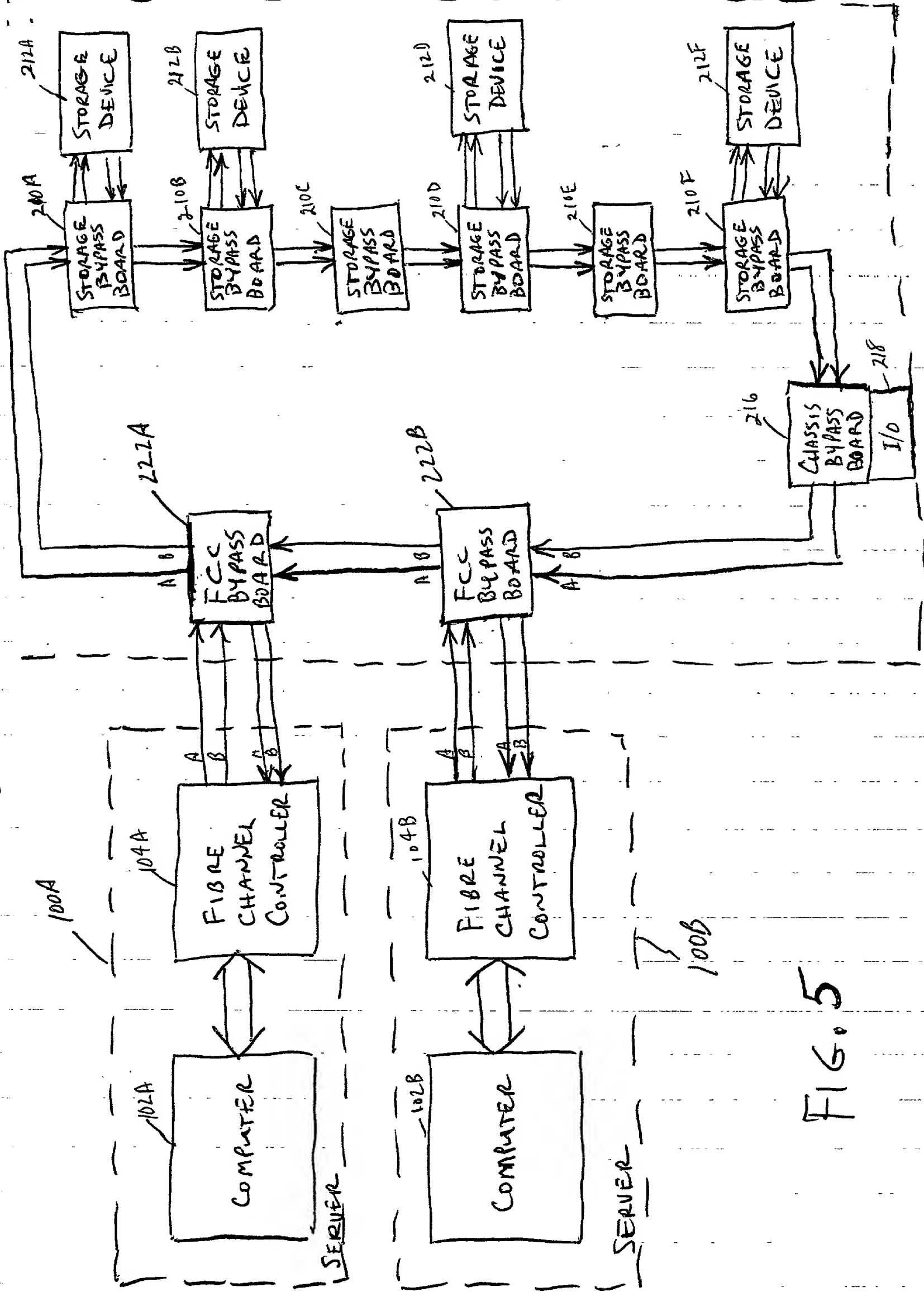
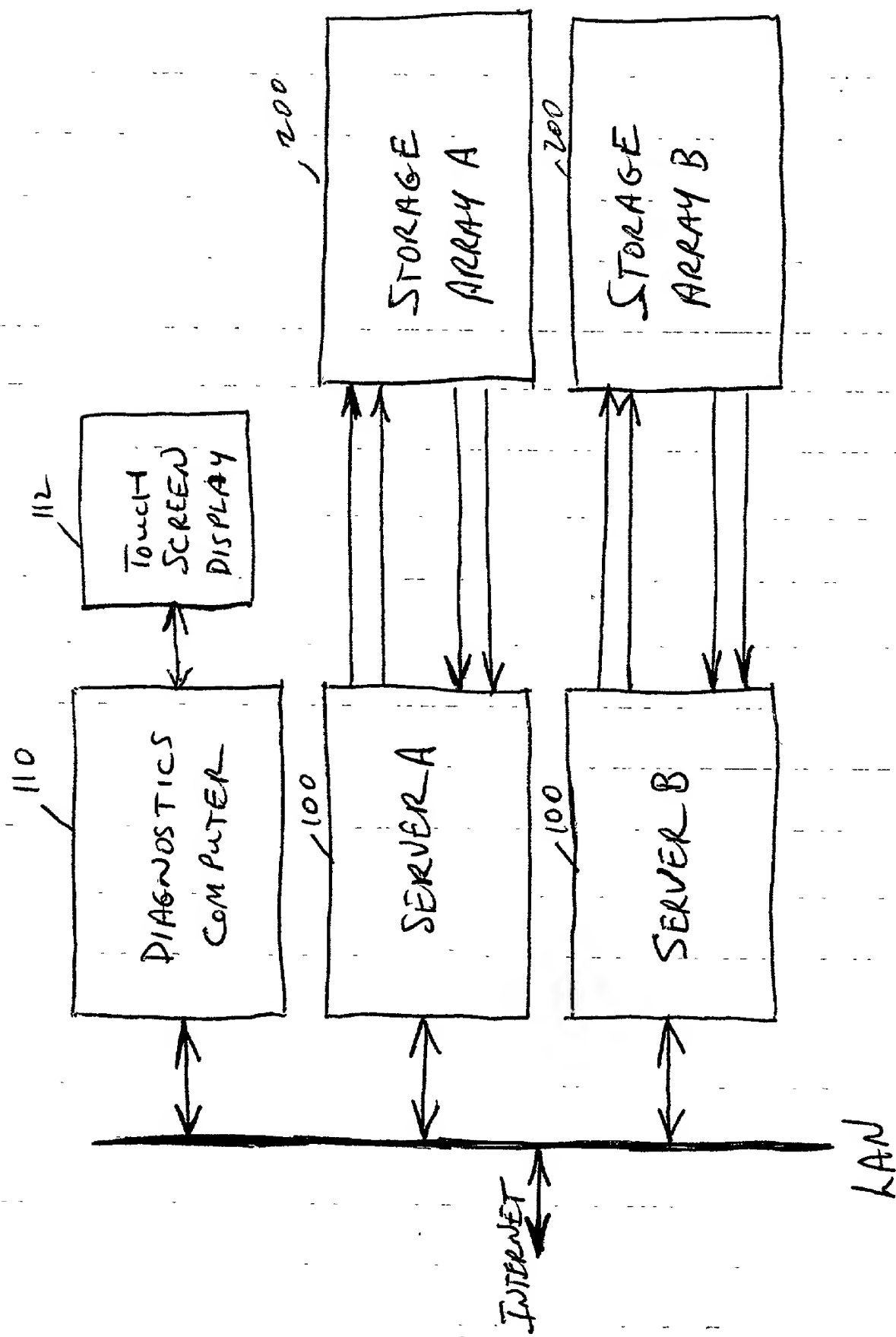


FIG. 5



462

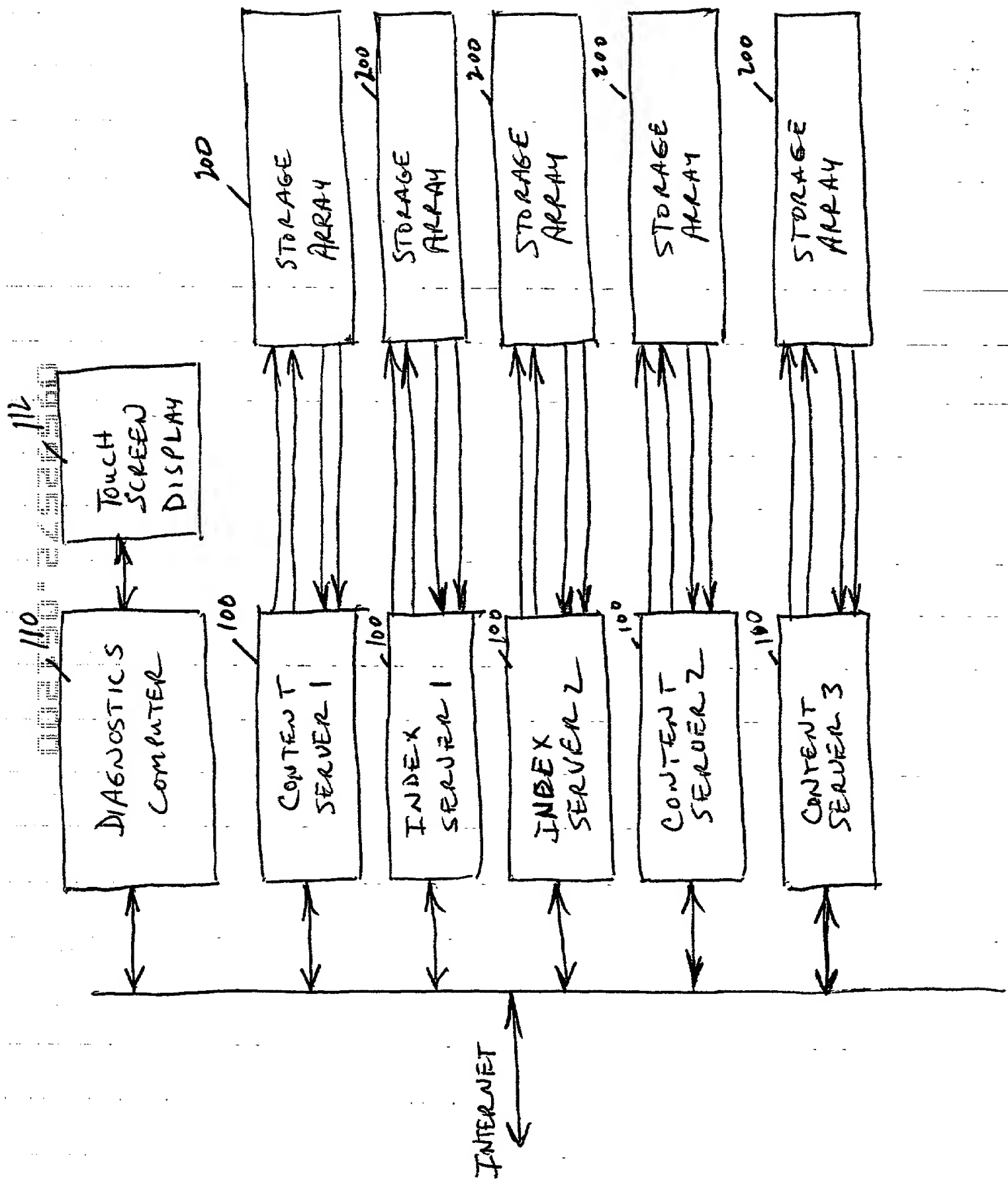


FIG. 8

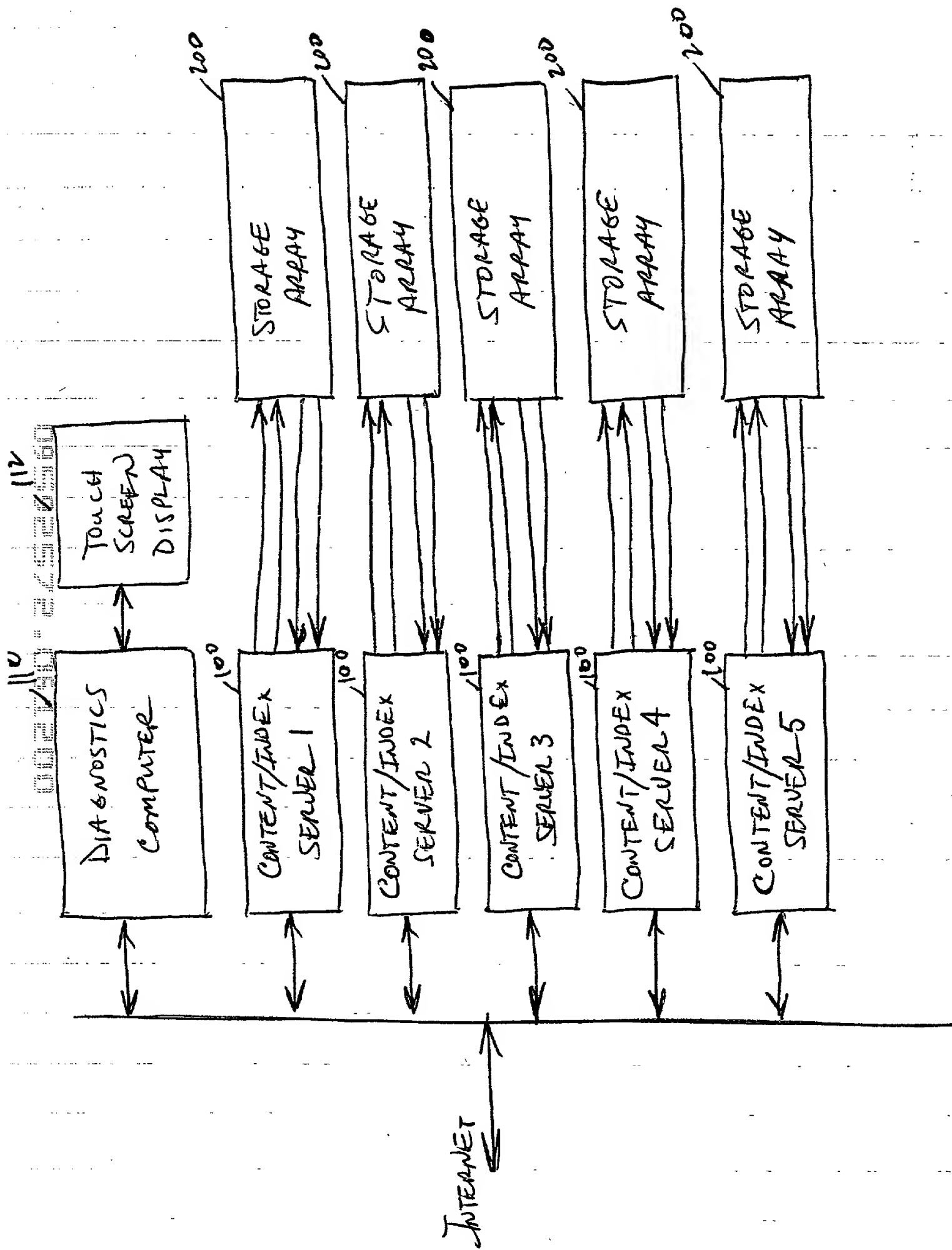


FIG. 9



100

216

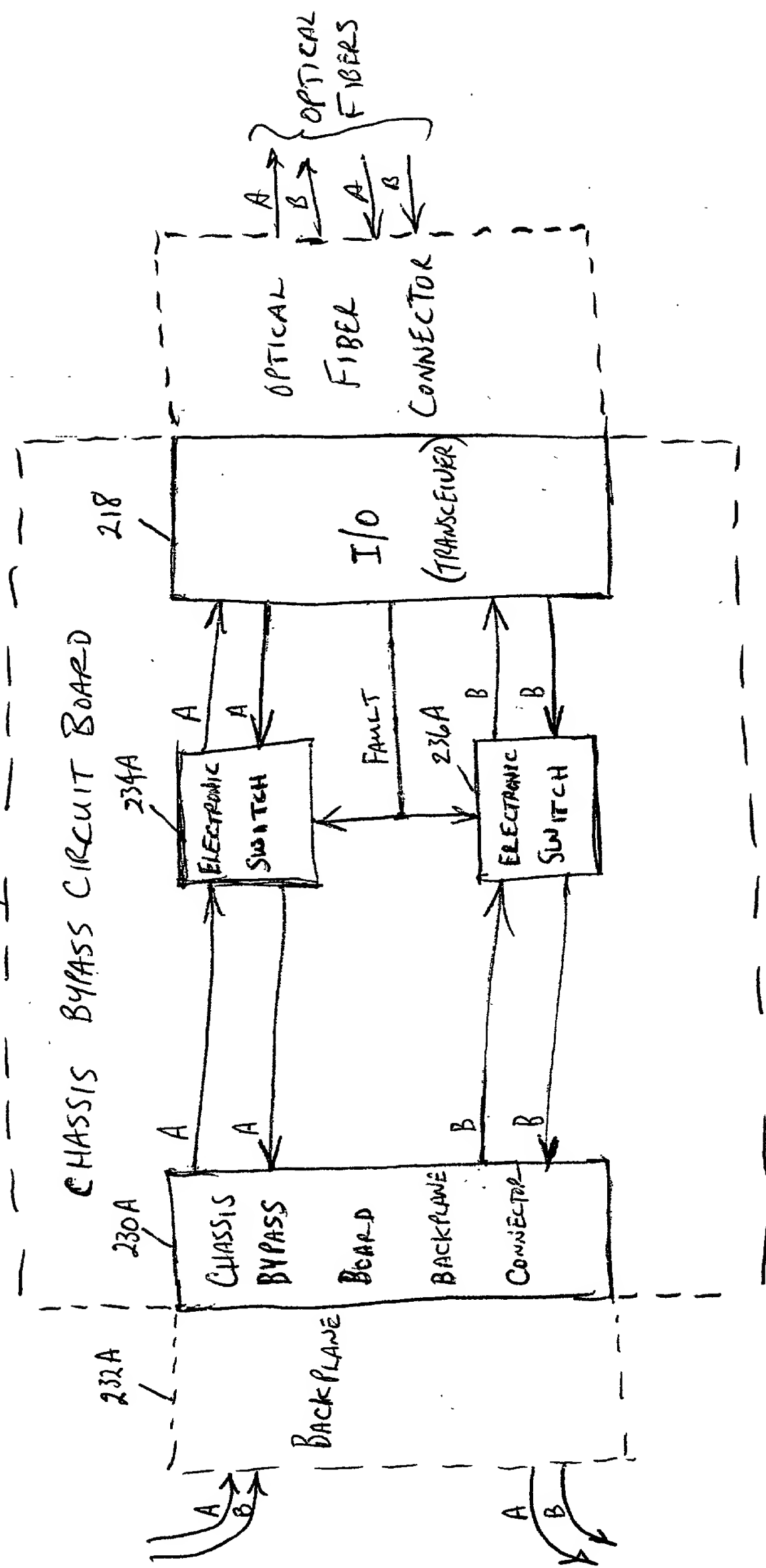


FIG. 11

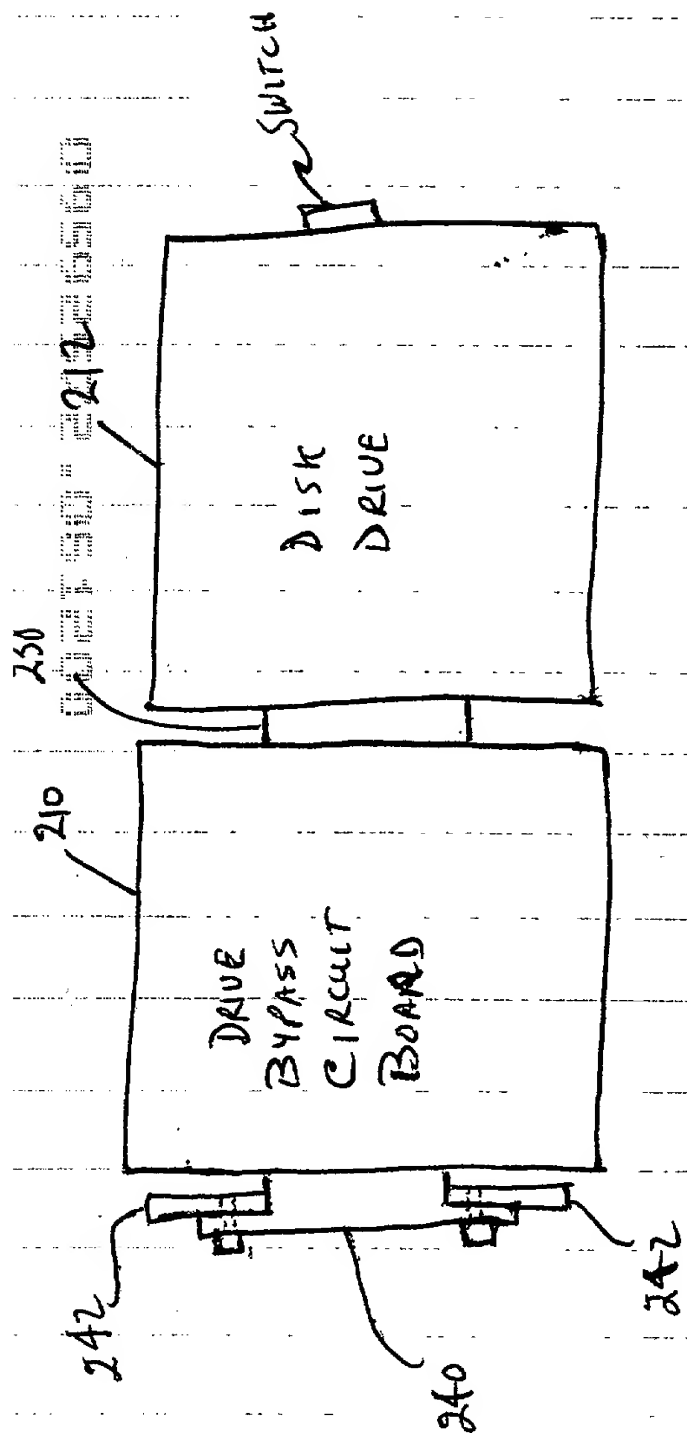


FIG. 12A

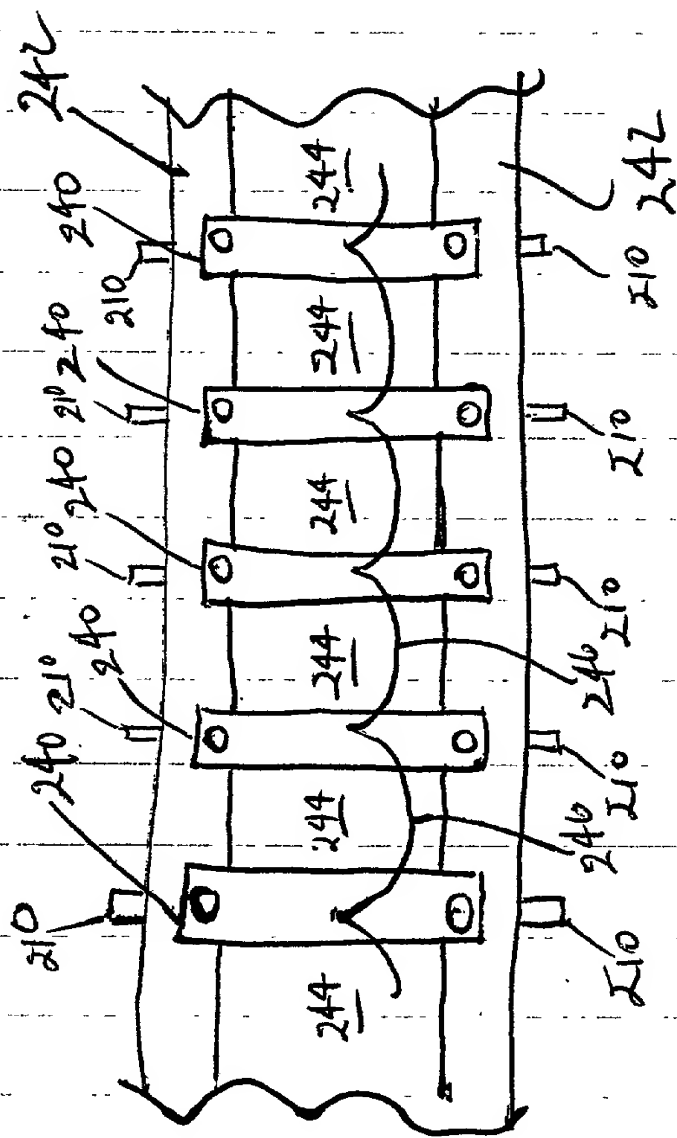


FIG. 12B

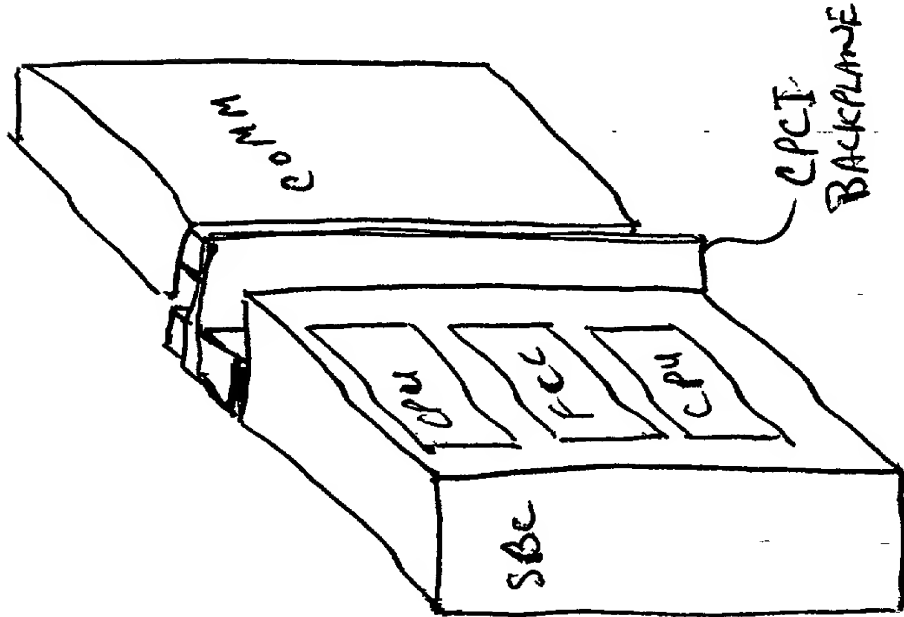


FIG. 13A

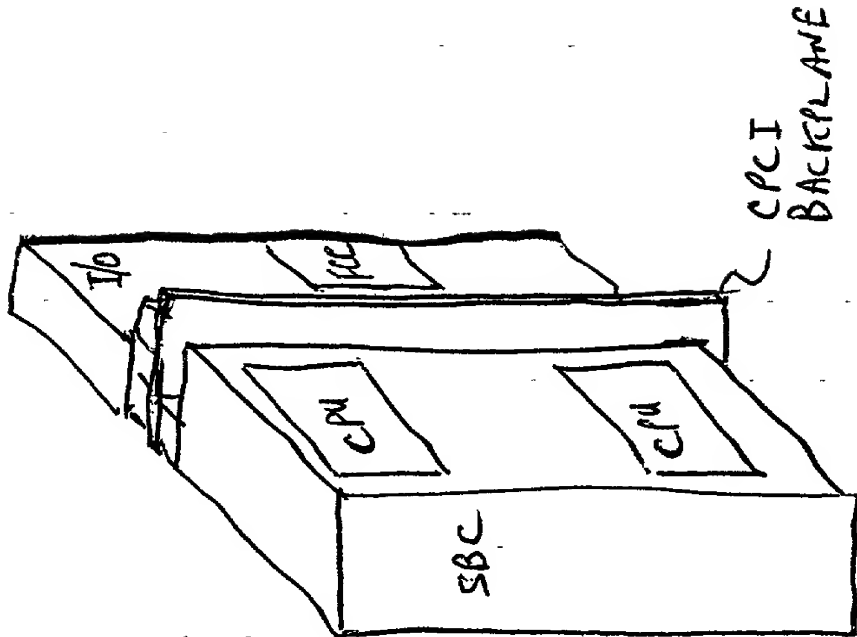


FIG. 13B

**COMBINED DECLARATION AND POWER OF ATTORNEY
FOR UTILITY PATENT APPLICATION**

Attorney's Docket No.

032933-001

As a below-named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name;

I BELIEVE I AM THE ORIGINAL, FIRST AND SOLE INVENTOR (if only one name is listed below) OR AN ORIGINAL, FIRST AND JOINT INVENTOR (if more than one name is listed below) OF THE SUBJECT MATTER WHICH IS CLAIMED AND FOR WHICH A PATENT IS SOUGHT ON THE INVENTION ENTITLED:

HIGH SPEED INFORMATION PROCESSING AND MASS STORAGE SYSTEM AND METHOD, PARTICULARLY
FOR INFORMATION AND APPLICATION SERVERS

the specification of which

(check one)

☐ is attached hereto;

☐ was filed on _____ as

Application No. _____

and was amended on _____;
(if applicable)

I HAVE REVIEWED AND UNDERSTAND THE CONTENTS OF THE ABOVE-IDENTIFIED SPECIFICATION, INCLUDING THE CLAIMS, AS AMENDED BY ANY AMENDMENT REFERRED TO ABOVE;

I ACKNOWLEDGE THE DUTY TO DISCLOSE TO THE OFFICE ALL INFORMATION KNOWN TO ME TO BE MATERIAL TO PATENTABILITY AS DEFINED IN TITLE 37, CODE OF FEDERAL REGULATIONS, Sec. 1.56 (as amended effective March 16, 1992);

I do not know and do not believe the said invention was ever known or used in the United States of America before my or our invention thereof, or patented or described in any printed publication in any country before my or our invention thereof or more than one year prior to said application; that said invention was not in public use or on sale in the United States of America more than one year prior to said application; that said invention has not been patented or made the subject of an inventor's certificate issued before the date of said application in any country foreign to the United States of America on any application filed by me or my legal representatives or assigns more than twelve months prior to said application;

I hereby claim foreign priority benefits under Title 35, United States Code Sec. 119 and/or Sec. 365 of any foreign application(s) for patent or inventor's certificate as indicated below and have also identified below any foreign application for patent or inventor's certificate on this invention having a filing date before that of the application(s) on which priority is claimed:

COMBINED DECLARATION AND POWER OF ATTORNEY

Attorney's Docket No.

032933-001

COUNTRY/INTERNATIONAL	APPLICATION NUMBER	DATE OF FILING (day, month, year)	PRIORITY CLAIMED
			YES_ NO_
			YES_ NO_

I hereby appoint the following attorneys and agent(s) to prosecute said application and to transact all business in the Patent and Trademark Office connected therewith and to file, prosecute and to transact all business in connection with international applications directed to said invention:

William L. Mathis	17,337	R. Danny Huntington	27,903	Gerald F. Swiss	30,113
Robert S. Swecker	19,885	Eric H. Weisblatt	30,505	Michael J. Ure	33,089
Platon N. Mandros	22,124	James W. Peterson	26,057	Charles F. Wieland III	33,096
Benton S. Duffett, Jr.	22,030	Teresa Stanek Rea	30,427	Bruce T. Wieder	33,815
Norman H. Stepno	22,716	Robert E. Krebs	25,885	Todd R. Walters	34,040
Ronald L. Grudziecki	24,970	William C. Rowland	30,888	Ronni S. Jillions	31,979
Frederick G. Michaud, Jr.	26,003	T. Gene Dillahunt	25,423	Harold R. Brown III	36,341
Alan E. Kopecki	25,813	Patrick C. Keane	32,858	Allen R. Baum	36,086
Regis E. Slutter	26,999	Bruce J. Boggs, Jr.	32,344	Steven M. du Bois	35,023
Samuel C. Miller, III	27,360	William H. Benz	25,952	Brian P. O'Shaughnessy	32,747
Robert G. Mukai	28,531	Peter K. Skiff	31,917	Kenneth B. Leffler	36,075
George A. Hovanec, Jr.	28,223	Richard J. McGrath	29,195	Fred W. Hathaway	32,236
James A. LaBarre	28,632	Matthew L. Schneider	32,814		
E. Joseph Gess	28,510	Michael G. Savage	32,596		



21839

and:

Address all correspondence to:



21839

Frederick G. Michaud, Jr.
BURNS, DOANE, SWECKER & MATHIS, L.L.P.
P.O. Box 1404
Alexandria, Virginia 22313-1404

Address all telephone calls to: Frederick G. Michaud, Jr. at (703) 836-6620.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

FULL NAME OF SOLE OR FIRST INVENTOR	SIGNATURE	DATE
Richard Dellacona		
RESIDENCE	CITIZENSHIP	
6105 Orchard Grove Way, Riverside, CA 92505	US	
POST OFFICE ADDRESS		
6105 Orchard Grove Way, Riverside, CA 92505		
FULL NAME OF SECOND JOINT INVENTOR, IF ANY	SIGNATURE	DATE
RESIDENCE	CITIZENSHIP	
POST OFFICE ADDRESS		
FULL NAME OF THIRD JOINT INVENTOR, IF ANY	SIGNATURE	DATE